利用卷积神经网络改进迭代深度学习算法的图像识别方法研究 *

孙平安 [1a]，祁　俊 [2]，谭秋月 [1b]

(1. 武夷学院 a. 实验室管理中心; b. 数学与计算机学院, 福建 武夷山 354300; 2. 青海大学 计算机技术与应用系, 西宁 810016)

摘　要：针对图像识别算法中图像集上几何曲面的特定分类会导致判别信息丢失的问题，提出一种融合卷积神经网络的改进型迭代深度学习算法(IIDLA)。该算法采用混合卷积网(PCL)进行底层的平移不变特征学习，以层次化的方式迭代应用卷积神经网络(CNN)对输入图像集的不同非线性特征进行学习。算法的图库和查询实例中包括了不同视角、背景、面部表情、解析度和照明度的人脸或物体图像集。采用数据集将提出的算法与其他算法进行评估对比，实验结果表明，提出的算法在被测数据集上的性能最优。

关键词：深度学习；卷积神经网络；自适应；图像识别；层次化迭代

中图分类号：TP391.4　　doi: 10.3969/j.issn.1001-3695.2018.01.0071

Research of image recognition based on improved iterative depth learning algorithm using Convolutional Neural Network

Sun Ping'an[1a], Qi Jun[2], Tan Qiuyue[1b]

(1. *a. Laboratory Management Center*, *b. College of Mathematics & Computer*, *Wuyi University*, *Wuyishan Fujian* 354300, *China*; 2. *Dept. of Computer Technology & Application Qinghai University*, *Xining* 810016, *China*)

Abstract: Aiming at the problem that the specific classification of geometric surfaces on the set of image recognition algorithms can lead to the loss of discriminative information, this paper proposes an Improved Iterative Depth Learning Algorithm (IIDLA) based on convolution neural network. The algorithm uses the hybrid convolution network (PCL) to perform the translation invariant feature learning at the bottom, and uses iteratively applies convolutional neural network (CNN) to learn different nonlinear characteristics of the input image set in a hierarchical way. The graphic library and query examples of the algorithm include different perspectives, backgrounds, facial expressions, resolution and illumination, and the image sets of objects or objects. A data set is used to compare the proposed algorithm with other algorithms, and the experimental results show that the proposed algorithm has the best performance on the measured data set.

Key words: Depth learning; Convolutional Neural Network; Adaptive; Image recognition; Layering iterative

0　引言

面部和物体识别在现实生活有着各种各样的应用，如人机交互、监控、普适计算和访问控制等[1]。光照和视角变化带来的影响使得人脸/物体识别成为一项非常具有挑战性的任务[2,3]。近年来，基于图像集的人脸/物体识别备受关注[4]，这主要是由于以下两个原因：首先，相比于单一人脸照片，图像集提供了更多关于目标面部或其他外表的信息。在面部表情、照明条件、环境、视角等条件变化的环境下，一段时间内捕获的人脸/物体图像集均可供学习、训练使用 [5]。其次，低成本便携式传感器的广泛使用使得获取图像集更为便捷，如监控场景、室内机器人导航和定位等。虽然基于图像集的人脸/物体识别具备较好的分类性能，但是对急剧增加的庞大数据量进行有效处理并识别仍然亟待解决。简而言之，基于图像集的针对面部表情、光照、环境、视角变化条件下人脸/物体的识别仍然是研究的热点。

文献[6]在非线性流形上对图像集的变化信息进行几何建模，文献[7]在格拉斯曼流形上对图像集的变化进行几何建模；而文献[8,9]分别在黎曼流形、子空间模拟上对图像集进行几何建模。这三种算法均会对图像集几何表面上的特定类做出预先假设，导致部分判别信息丢失，而判别信息的丢失所引起的识

收稿日期：2018-01-25；**修回日期：**2018-04-02　　**基金项目：**国家自然科学基金资助项目（61762086）；福建省教育厅信息化类项目（JAT160524）；福建省教育厅科技 A 类项目（JA15513）

作者简介：孙平安（1977-），男，陕西渭南人，副教授，硕士，主要研究方向为智能图像处理、机器学习等（pingan171115@126.com）；祁俊（1977-），女，青海贵德人，副教授，硕士，主要研究方向为图像处理，智能信息处理等；谭秋月（1980-），女，陕西咸阳人，副教授，硕士，主要研究方向为图像处理等.

别偏差将不可避免。对于图像集分类所面临的如何建模获得图库集之间的距离，即相似度的问题也有较多研究。文献[9]采用参数建模法，以参数分布的形式表示图像集，使用相对熵来度量分布之间的相似性，但算法需要图库和查询图像集之间有较强的统计关系才能实现较好的性能。如果测试图像集和训练图像集之间没有较强的统计关系，将无法得到有效的参数，而且算法对数据集规模大小也有一定的要求，即过小规模的数据集将导致该算法失效。

为了改善参数化建模法的不足，非参数建模法越来越受关注，与参数化建模法不同的是非参数建模法不依赖于数据分布规律的假设。文献[10]提出了图像集分类的非参数建模法，算法通过典型样本逼近图像集，性能表现达到了预期效果。基于几何表面的算法利用几何表面上的点表示完整的图像，图像集可以在复杂的非线性流形上建模，也可以利用表面或混合表面建模[11]。文献[12]从几何图像的仿射包或凸包算法中得到的集合样本，使用特征角来确定以子空间作为表征的图像集间距离。两个子空间的特征角 $0 \le \theta_1 \le \cdots \theta_d \le \pi/2$ 定义为一个子空间中任意向量和第二个子空间中任意向量之间夹角的最小值，而主角的余弦总和用于计算子空间之间的相似度。此算法优点是不用担心异常值的影响且计算开销不大，但是在图像集样本规模较小且变化很大时性能降低严重。文献[13]使用图像集的仿射包或凸包算法来学习集合样本，对于以典型实例表示的图像集，将集合-集合距离定义为集合代表之间的欧氏距离，以简单地自适应学习集合样本。文献[14]提出了稀疏近似点(sparse approximated nearest points，SANP)算法，首先从相应集合的均值图像和仿射包图像算法计算两组集合的 SANP。然后，通过 SANP 稀疏近似集合的样本图像来寻找集合之间的最近点。但是文献[13][14]的算法都需要计算一对一映射的集合距离，对计算开销较大。近年来有文献提出一些改进的学习算法，如基于流形-流形距离算法(manifold-manifold distance algorithm，MMDA)、基于仿射包图像集距离算法(distance algorithm based on affine packet image set，DAPIS)、基于凸包的图像集距离算法(image set distance algorithm based on convex hull，ISDACH)等。文献[15]提出了基于图像集分类的自适应深度网络算法(adaptive deep network，ADNT)，深度学习的框架由编码层和解码层组成，编码层和解码层用于重建输入图像。ANDT 使用手动生成的局部二值模式（local binary pattern，LBP）特征进行训练和测试，有助于获得良好的性能，但是增加了额外的计算复杂度。

因此，提出了一种改进型迭代深度学习算法(improved iterative depth learning algorithm，IIDLA)。算法通过非线性激活函数将单层卷积神经网络(convolutional neural network，cnn)、混合卷积网(pooled convolutional layer，PCL)和多重人工神经网络(Artificial Neural Networks，ANN)层有机融合，再经过层次化迭代避免信息丢失、计算代价高昂、小样本失效、异常值影响等问题。实验结果表明提出的算法在人脸识别、物体分类的性能上取得了较好的效果。

1　改进型迭代深度学习算法

深度学习模型发展至今已有卷积神经网络、深度置信网络 、层叠自动去噪编码机等不同模型。其中卷积神经网络通过卷积操作自动提取输入数据的不同特征，已经成为了无监督预训练神经网络的代表。

本文提出的 IIDLA 一定程度上借鉴了卷积神经网络的理论架构并进行改进。提出的 IIDLA 由编码层、解码层和隐藏层组成，需要注意的是 IIDLA 与文献[15]中的 ADNT 有着明显的差异。首先，在 ADNT 中是随机选择层数和每一层的大小。其次，ADNT 的性能取决于手动生成的 LBP 特征。第三，ADNT 的自编码基于稀疏编码的概念，而稀疏编码的最大缺点是是对于大的输入维数，在速度方面的伸缩性较差。第四，ADNT 需要非常谨慎地选择隐藏层的节点数。因为如果隐藏层的节点数比输入、输出层的少，则会导致算法将最终隐藏层的启动当成输入的压缩表示。而当隐藏层比输入层大时，基于 ADNT 的自编码将潜在地学习标识函数，从而导致算法效率低下。与 ADNT 算法的以上缺点不同，IIDLA 由相同的 ANNs 组成，它迭代地用于树结构中以学习分层特征表示。树的叶节点表示为 L 维向量，集中形成 3D 矩阵，并合并卷积层的输出，如图 1、图 2 所示。

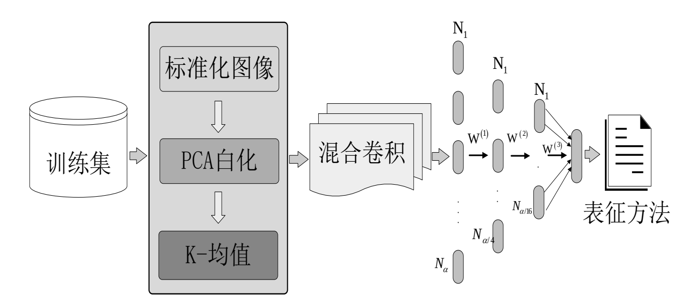


图 1 迭代深度学习算法(IIDLA)的训练示意图

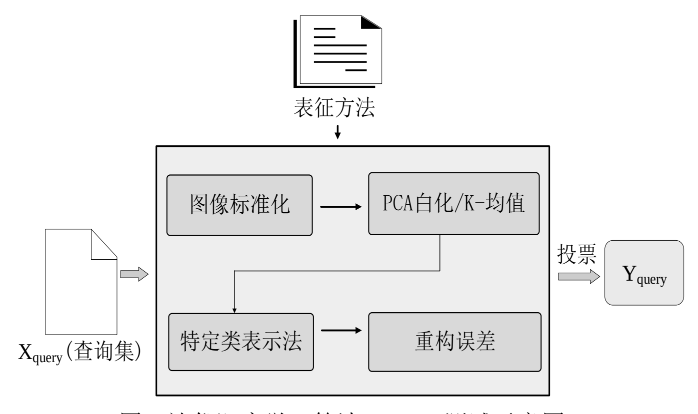


图 2 迭代深度学习算法(IIDLA)测试示意图

因此，层数和 ANNs(N_α)的数量取决于 X 的尺寸，IIDLA 具有自适应性且对于任意给定的 X ，IIDLA 都可以被调整。此外，提出的 IIDLA 没有任何隐藏层，压缩表示或陷入恒等函数学习的概率很小，与依赖于手动生成特征的 ADNT 相比，IIDLA

使用 PCL 从原始图像中自动地学习判别式表示法。PCL 的卷积滤波器提供了平移不变性特征，如边界，允许部分人脸/物体在一定程度上是可变的，经过设计对比可以预见 IIDLA 相比于 ADNT 会实现更优越的性能。

1.1　分类算法的训练算法

如图 3 描述了提出的训练算法：首先令 $\chi_1,\chi_2,\chi_3\cdots\chi_n$ 为 n 个图库图像集，且它们相应的标签为 $L_\varphi\in[1,2,\ldots,n]$，其中，图像集 $\chi_\varphi\in\left\{I^{(t)}\middle|L^{(t)}=\varphi;t=1,2,\ldots,K_\varphi\right\}$ 有属于类 φ 的 K_φ 图像 $I^{(t)}$。对于 d 个图像类中的每一个，将其经过 PCL 滤波器，并以非监督的方式初始化 ANNs 的权值。通过将来自所有训练图库集的人脸/物体图像集集合到数据集 $\chi_u=U\left\{\forall I^{(t)}\in X_\varphi;\forall\varphi\in[1,2,\ldots n]\right\}$ 中，生成非监督图库数据集。然后，使用从图库集 χ_i 随机采样的图像块的 k-均值进行 PCL 滤波的非监督训练。根据经验选择 16×16 的块，然后，通过减去均值和利用其元素的标准偏差进行划分来归一化。为了在原始图像中除去除冗余的特征，还实施了 PCL 白化。接下来，利用大小为 $f=16\times16$ 的滤波器和和大小为 $L=128$ 的滤波器组执行卷积。对于平均池化，步长 $s=5$；池化区域尺寸 $r=10$ 用于为每个图像生成尺寸为 $128\times64\times64$ 的 3D 矩阵。

提出的 IIDLA 用 d 个图库图像集中的所有图像集进行训练。且用权值 $W^{(a)},a=1,\ldots,3$ 初始化每个 IIDLA，使得 IIDLA 的每层都具有应用空间为尺寸 $\beta\times\beta$ 的子空间。根据经验，在本文中选择 $\beta=4$，于是在每层深度网络上运用下列矩阵：$X\in R^{128\times64\times64}$ 到 $p^{(1)}\in R^{128\times16\times16}$ 到 $p^{(2)}\in R^{128\times4\times4}$ 到最后的 $p\in R^{128}$。通过最小化遍历所有图库集 X_φ 实例上的重建误差反向传播，来执行随机梯度下降算法，从而完成类特定表征的学习。为了避免过度拟合，引入正则化项来调整以上成本函数，如调整权重衰减损失项、稀疏约束项等。然后通过训练遍历 X_φ 集的所有图像 3D 矩阵获得特定类的表示 δ_φ，如下：

$$\delta_\varphi=\min_{\delta_{IDLM}}J\left(\delta_{IDLM};X^{(t)}\in X_\varphi\right)\qquad(1)$$

因此，特定类的表示 δ_φ 用于学习输入图像的底层结构。由于所使用的激活函数是非线性的，所以提出的算法能够学习非常复杂的非线性结构。需要注意的是，基于此算法提出的 IIDLA 还是可扩展的，增加新的图像集不需要在完整的数据集上重新训练。

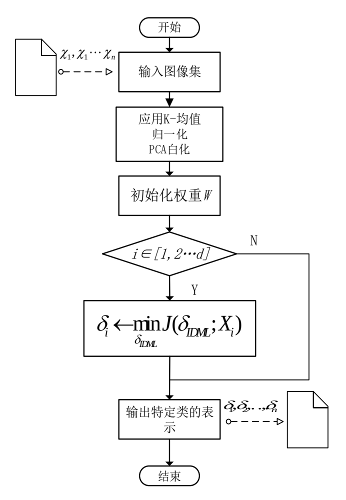


图 3　迭代深度学习算法的训练流程示意图

混合卷积网络的正反向传播过程如下：

a)正向传播。混合卷积神经网络通过激励-响应机制计算每个神经元的输入，进行正向传播，其每个神经元的输出记作：

$$x_n^i=f(y_n^i)=f\left(\sum_{j=0}^{C_{n-1}}\omega_n^{ji}x_{n-1}^j\right)\qquad(2)$$

其中：x_n^i 为第 n 层的第 i 个神经元的输出；ω_n^{ji} 为第 n-1 层的第 j 个神经元到第 n 层的第 i 个神经元的权重；C_{n-1} 为第 n-1 层神经元个数；

b)反向传播。对于由一系列确定的单元互连形成的混合卷积网络，反向传播用于学习网络的权值。

$$\Delta\omega_n^{ji}=-\eta\frac{\partial E_n}{\partial\omega_n^{ji}}\qquad(3)$$

其中 $$E_n=\frac{1}{2}\sum_{i=1}^{C_n}(x_n^i-T_n^i)^2\qquad(4)$$

其中：$\Delta\omega_n^{ji}$ 表示每次反向传播后权值的变化量；E_n 表示第 n 层的输出误差；C_n 表示第 n 层神经元个数；T_n^i 表示第 n 层的期望输出结果。

式中的 $\frac{\partial E_n}{\partial\omega_n^{ji}}$ 可以通过下式进行推导：

$$\frac{\partial E_n}{\partial\omega_n^{ji}}=\frac{\partial E_n}{\partial y_n^i}\frac{\partial y_n^i}{\partial\omega_n^{ji}}=x_{n-1}^j\frac{\partial E_n}{\partial\omega_n^{ji}}\qquad(5)$$

$$\frac{\partial E_n}{\partial y_n^i}=G(x_n^i)\frac{\partial E_n}{\partial x_n^i}\qquad(6)$$

$$\frac{\partial E_n}{\partial x_n^i}=x_n^i-T_n^i\qquad(7)$$

计算出第 n 层权值 ω_n^{ji} 的增加值。而第 n-1 层 $\frac{\partial E_{n-1}}{\partial x_{n-1}^i}$ 计算

式为

$$\frac{\partial E_{n-1}}{\partial x_{n-1}^{j}} = \sum \omega_n^{kj} \frac{\partial E_n}{\partial y_n^{i}} \tag{8}$$

按以上公式类推计算每一层，可以得到：

$$\omega_{n-now} = \omega_{n-pre} - \eta \frac{\partial E_n}{\omega_n} \tag{9}$$

其中：ω_{n-now} 表示更新后的权值；ω_{n-pre} 表示更新前的权值。

1.2　基于图像集的分类算法

图 4 描述了提出的基于图像集的分类算法流程。考虑到 n 个图库图像集和其标签 L_φ，基于分类的图像集问题转化为：给定查询集 $\chi_{query} = \{I^{(t)}\}_{1\times K_{query}}$，查找 χ_{query} 的类 L_{query}。对于所给的 χ_{query}，分别重建来自所有特定类表示 $\delta_\varphi, \varphi = 1,2,\dots,n$ 的每种图像集 $I^{(t)} \in \chi_{query}$。如果 $\tilde{X}^{(t)}(\varphi)$ 是来自表示 δ_φ 的图像 $X^{(t)}$ 的重建，那么重建误差为：

$$\varepsilon^{(t)}(\varphi) = \left\| x^{(t)} - \tilde{X}^{(t)}(\varphi) \right\|_2 \tag{10}$$

一旦算出了所有 n 个表示的重建误差，对每个 $I^{(t)} \in \chi_{query}$ 进行投票。投票总数 $L^{(t)}$ 为投给利用最小重建误差重建图像的类，如 $\varepsilon^{(t)}(\varphi)$。接着计算由所有 χ_{query} 图像所投的票数，同时确定获得最高票数的候选类为查询图像集的类 L_{query}。

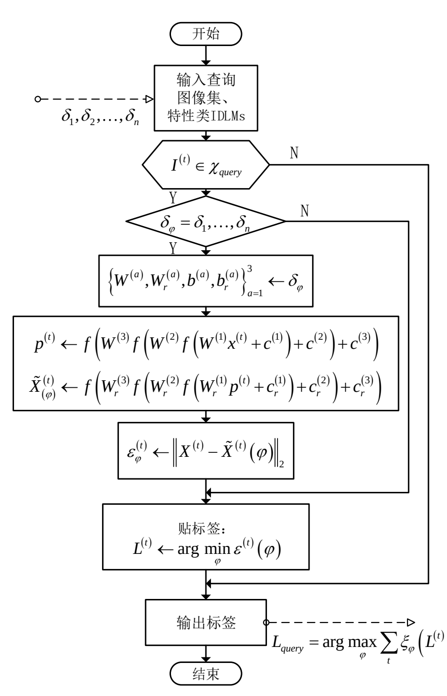


图 4 IIDLA 的图像集分类算法流程示意图

1.3　IIDML 的非监督预训练

虽然提出的深度学习网络已经将图像像素值用作原始特征，但如果想要获得更好的性能，适当的预训练还是必要的。IIDLA 的预训练流程如下，IIDLA 的非监督预训练分为以下两个阶段。

在第一个阶段，以非监督的方式初始化 PCL 滤波器。首先，从给定的图像集中提取随机块，然后归一化和白化这些块。接下来，应用 k-均值算法聚类预处理块：

$$\arg\min_{I} \sum_{i=1}^{z} \sum_{p\in S_i} \left\| p - \eta_i \right\| \tag{11}$$

其中：μ_i 是给定图像集 S_i 中点的均值。需要注意的是，对于预训练，使用 k-均值算法有三个主要优点：计算非常高效；实施简单；没有超越算法结构自身的超参数。

在第二阶段，通过使用 GRBMs 实施非监督分层预训练来初始化 ANNs 的权值。其中 GRBM 是标准鲁棒玻尔兹曼机（robust Boltzmann machine，RBM）的一个应用较为广泛的扩展延伸。其概率分布如下式：

$$p\left(h_j = 1 | v\right) = sigmoid\left(\sum_i \omega_{ij} v_i + f_j\right) \tag{12}$$

$$p\left(v_i | h\right) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(\frac{-\left(v_i - d_i\right)^2}{2\sigma_i^2}\right) \tag{13}$$

其中：$d_i = e_i + \sigma_i^2 \sum_j \omega_{ij} h_j$　(14)

σ_i 是到可见节点 v_i 的真实高斯分布式输入的标准偏差，h_j 为隐藏节点，e 和 f 分别为可见层和隐藏层的偏差。

1.4　IIDLA 的表征

IIDLA 初始化后，为了提取不变特征，将以非监督方式学到的 PCL 参数/滤波器卷积到输入原始图像中。用大小 $k\times k$ 的 L 方形滤波器卷积大小 $N\times N$ 的每个输入图像，结果产生 L 滤波器响应，每个响应滤波器大小为 $(N-k+1)\times(N-k+1)$。合并的卷积层使用它的非线性特性，校正和局部对比度归一化之后即为卷积。接着，将大小为 $(N-k+1)\times(N-k+1)$ 的滤波器的响应输出和大小为 $j\times j$、跨步为 s 的正方形区域合并，从而获得宽度和高度的混合响应，如下：

$$\alpha = N - j / s + 1 \tag{15}$$

对于每个输入的原始图像，PCL 的输出为大小为 $L\times\alpha\times\alpha$ 的 3D 矩阵 X。对于所给的 3D 矩阵 X，定义了由 X 中邻近向量组成的 $L\times\beta\times\beta$ 尺寸块。需要注意的是，在水平和垂直方向已经使用了 4 个邻近向量；所以在这个情况下 $\beta = 4$。因此，得到的块大小为 $L\times 4\times 4 (L = 128)$，它由 16

个向量组成。另外，由于 $\beta=4$，有 $\beta^2=16$ 个向量。随后，通过将输入 X 映射到父向量 p，从而将向量合并到 p 中，如下式：

$$p^{(1)}=f\left(W^{(1)}X+c^{(1)}\right) \quad (16)$$

$$p^{(2)}=f\left(W^{(2)}p^{(1)}+c^{(2)}\right) \quad (17)$$

$$p=f\left(W^{(3)}p^{(2)}+c^{(3)}\right) \quad (18)$$

其中：$W^{(a)}\in R^{L\times\beta^2L}$ 是参数矩阵，$\alpha=1,\ldots,3$，$f(\cdot)$ 为非线性激活函数(主要为 S 形函数或正切双曲线)，$c^{(i)}$ 为偏移向量，$p^{(1)}$、$p^{(2)}$ 和 p 分别是维度 $R^{L\times\alpha/4\times\alpha/4}$、$R^{L\times\alpha/16\times\alpha/16}$ 和 R^L 的矩阵。为了评估 L 维向量 p 对子矩阵的表征情况，使用父向量 p 进行重构。IIDLA 的重构结果给出如下：

$$x_1=f\left(W_r^{(1)}p+c_r^{(1)}\right) \quad (19)$$

$$x_2=f\left(W_r^{(2)}p+c_r^{(2)}\right) \quad (20)$$

$$x_3=f\left(W_r^{(3)}p+c_r^{(3)}\right) \quad (21)$$

其中：$W_r^{(a)},a=1,\ldots,3$ 为由 $W_r^{(1)}=W^{(3)^T}$、$W_r^{(2)}=W^{(2)^T}$ 和 $W_r^{(3)}=W^{(1)^T}$ 所给出的权值，$c_r^{(i)}$ 为偏移向量。因此，可通过 IIDLA 的参数 $\delta_{IDLM}=\left\{\delta_W,\delta_c\right\}$ 来表示完整的 IIDLA，其中 $\delta_W=\left\{W^{(a)},W_r^{(i)}\right\}$，$\delta_b=\left\{c^{(i)},c_r^{(i)}\right\}$。然后，针对图库图像集的所有类，分别训练 IIDLA。

2 仿真验证及分析

本算法是针对图像集任务的人脸识别和物体分类提出的，对于人脸识别的性能评估在 DMIY、YT 和 FOCUT 数据集上完成，对于物体分类采用的数据集是应用较为普遍的 ETC-80 数据集，采用平均识别率和标准偏差作为评估的性能指标。将提出的算法与 MMDA、DAPIS、ISDACH、RNP、CDLA 和 ADNT 等几种算法进行对比分析。实验重复 20 次，每次实验都是随机选择图库和查询集以保证验证的有效性。下面将给出这些数据集的详细描述、实验结构以及新提出的算法与其他先进的图像集分类算法比较过程和比较结果。

2.1 实验环境搭建

实验中面部识别评估采用常用的数据集，如 FOCUT、YT、DMIY 和 ETC-80 数据集[16]。

FOCUT 数据集包含来自 295 个帧，像素尺寸为 720×480 的记录个体 510 个正面和 706 个非正面的视频序列。 序列是在不同的时间获得的,在这些序列中，个体站立并与摄像机的非正面人脸交谈。序列包含在会话期间发生的正常头部运动，例如头旋转 90 度、举手、指向某处等。FOCUT 使用著名的 VJ 人脸检测算法从视频帧中提取面部[17]。然后将所提取的面部图像大小调整为 128×128。在本实验中，将每个视频当作一个图像集，将 20 个视频序列用作图库集，剩下的 30 个作为查询集。FOCUT 数据集样本如图 5 所示。

YT 是用于基于面部识别的图像的最全面且最具挑战性的数据集。该数据集包含近 2000 个视频，这些视频是从网站上收集的 100 位名人的视频。该数据集的面部图像以光照、表情和姿势的形式展示了很大区别和变化。此外，由于图像的高压缩率，图像的质量和分辨率非常低。采用网站数据集是由于其可以在每个视频序列中跟踪人脸区域的优异特性由数据集提供的在第一帧中画面中的面部的位置用以开始跟踪。成功检测之后，截取面部区域，同时将所有图像转换为灰度级。然后将剪取的灰度图像调至 128×128。对于性能评估，采用 5 层交叉验证，在每层中每个对象有 9 个图像集，随机选择这些图像集中的三个作为图库集，而剩下的六个用作查询集。YT 数据集样本如图 6 所示。

DMIY 数据集包括 43 个对象的序列，每个对象包含在不同时刻记录的 3 个视频序列。平均帧数约为 100，分辨率为 512x384 像素。每个主题的面部图像的平均数量是 300。在每个视频中，人将头部向左，向右，向中心，向上，然后向下，最后再向中心移动。首先提取来自视频的面部，并将其大小调整为 128×128。将每个视频当成一个图像集，随机选择对象的一个序列作为图库集，剩下的 3 个序列用作查询集。DMIY 数据集样本如图 7 所示。

对于对象分类，使用 ETC-80 数据集。这个数据集包含八类对象，包括苹果、汽车、杯子、狗、马、梨和西红柿。每类对象包含 10 个子类别，如牛的不同品种或苹果的不同类型，每个子类有 41 个不同视角的图像。实验使用 128×128 大小的图像，将子类中对象的图像当成图像集，对于每个对象，选择五个子类作为图库集，剩下的五个用作查询集。ETC-80 数据集的八类对象样本如图 8 所示



图 5 FOCUT 数据集的样本图像



图 6 YT 数据集的样本图像



图 7 DMIY 数据集的样本图像

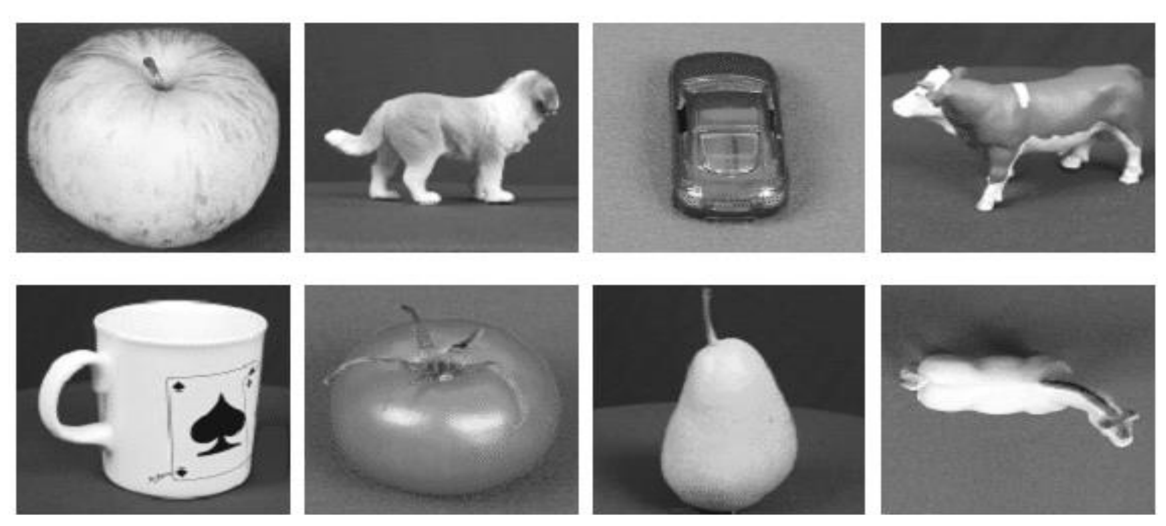


图 8 DMIY 数据集的样本图像

2.2 图像集分类算法的比较

表 1 总结了提出算法和领域内其他算法的平均识别率和标准偏差的数据对比。为了更直观的观察不同识别算法的差异，做出四种数据集下的综合平均识别率直方图如图 9 所示。综合表 1 和图 9 可以直观地看出，提出的算法获得了最优性能。由表 1 可以看到，提出的基于 IIDLA 的算法获得了最高识别率，FOCUT、YT、DMIY 和 ETC-80 数据集上获得的识别率分别是 77.52%、100%、98.53%和 98.72%。使用手动生成的特征可以进一步增加提出的 IIDLA 算法的性能。可以看到相比于 FOCUT 和 DMIY 数据集，YT 数据集所有算法获得的识别率很低。这是由于 YT 数据集最具有挑战难度，YT 数据集的视频是在外貌条件变化的真实生活场景中捕获的。

表 1 各算法在评估数据集上的平均识别率和标准偏差

算法	人脸识别			对象分类
	YT	FOCUT	DMIY	ETC-80
MMDA	54.32 ± 2.69	91.55 ± 2.56	91.21 ± 2.45	78.70 ± 4.80
DAPIS	62.29 ± 2.24	90.98 ± 2.18	93.09 ± 2.01	79.15 ± 5.20
ISDACH	61.02 ± 5.29	94.02 ± 1.83	96.85 ± 1.29	79.83 ± 5.12
GEDA	51.70 ± 4.39	91.39 ± 5.64	84.66 ± 3.04	79.42 ± 5.18
CDLA	55.89 ± 4.89	98.89 ± 1.52	90.11 ± 4.22	78.09 ± 4.19
MSSRC	58.99 ± 5.59	97.86 ± 2.78	97.68 ± 0.92	90.95 ± 3.19
RNP	66.17 ± 5.10	95.71 ± 2.35	95.79 ± 1.35	96.12 ± 2.71
ADNT	70.89 ± 4.98	100.0 ± 0.01	97.89 ± 0.32	98.19 ± 2.05
IIDLA	77.52 ± 4.02	100.0 ± 0.0	98.53 ± 0.39	98.72 ± 1.02

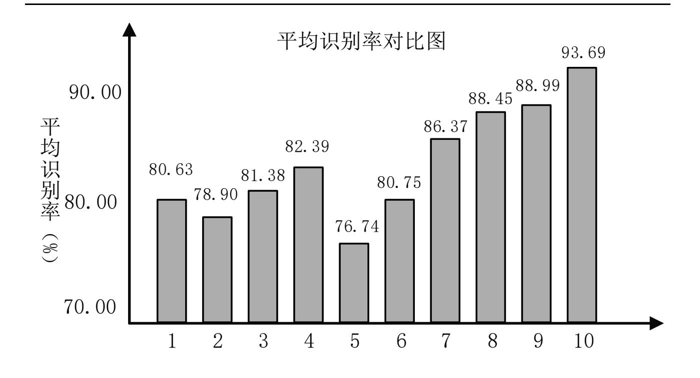


图 9 综合平均识别率对比图

2.3 图像集大小对 IIDLA 性能的影响

为了评估图像集样本数减少情况下 IIDLA 算法的性能，利

用较少的图像样本定义图库和查询集。在具有不同配置的图库和查询集的 FOCUT 上实施实验。使用的图库集尺寸为{*All*,100,70,35}，使用的查询集尺寸为{*All*,100,70,15,1}。表 2 为不同尺寸图库集和查询集的实验结果。表的第一行表示集中使用的图像样本的数量。例如，70–35 意味着图库集和查询集中分别使用 70 和 35 个图像样本。该表反应了不同算法的平均识别率和平均偏差，最后一列的"－"表示这些算法不能评估图库集中的一个图像样本。

结果表明新提出的迭代深度学习算法的性能优于其他所有对比算法的性能。需要注意的是甚至在图像集中的图像样本数非常少，甚至于仅仅为 15 的时候 IIDLA 仍然能够保持优异性能。当图像样本数进一步降低时，其性能会一定程度的降低。尽管如此，相比于其他算法 IIDLA 受图像集大小减少的影响依然较小，表明本算法对过拟合问题不敏感。如前所述，使用权重衰减损失和稀疏约束调整成本函数通过修正成本函数很容易解决数据集的规模不能太小的问题。测试结果显示，对于集中图像样本数量较少的情况，其他算法的性能会明显降低。然而与之相比基于凸包和仿射包的算法，如 ISDACH、DAPIS 和 RNP 等算法在不同图像规模上都获得了较好的性能。在仅有一个图像样本的图像集下，所有算法的性能都会显著降低，而本算法则在这种极端苛刻的条件下仍能获得较高的平均识别率。

表 2　图像集的大小对 IIDLA 性能的影响,

算法	All-All	100-100	70-70	70-35	70-15	35-70	All-1
MMDA	91.89 ± 2.32	85.46 ± 1.96	82.91 ± 4.58	83.44 ± 4.19	82.37 ± 3.65	81.67 ± 3.51	70.15 ± 2.85
DAPIS	91.05 ± 1.45	91.85 ± 3.19	89.99 ± 2.39	91.29 ± 3.43	90.21 ± 1.98	88.94 ± 1.95	77.59 ± 7.72
ISDACH	93.98 ± 1.56	90.89 ± 2.26	90.82 ± 2.19	90.01 ± 4.16	87.23 ± 2.53	86.35 ± 2.46	75.68 ± 8.65
CDLA	98.88 ± 1.89	95.93 ± 2.00	93.72 ± 2.16	91.54 ± 3.19	88.51 ± 4.62	89.22 ± 5.14	70.36 ± 5.81
MSSRC	97.90 ± 3.16	97.27 ± 1.76	94.22 ± 1.36	91.26 ± 3.15	91.18 ± 2.56	91.37 ± 2.29	73.56 ± 4.91
RNP	96.13 ± 2.35	92.56 ± 3.45	90.53 ± 3.16	89.64 ± 5.61	85.92 ± 2.86	85.68 ± 2.52	40.02 ± 8.16
ADNT	99.65 ± 0.16	98.89 ± 1.85	96.88 ± 2.89	92.87 ± 4.27	84.14 ± 4.22	84.04 ± 4.75	74.65 ± 7.49
IIDLA	100.00 ± 0.0	100.0 ± 0.0	97.96 ± 2.56	95.85 ± 2.02	84.31 ± 4.12	84.91 ± 4.65	75.16 ± 5.24

3　结束语

基于图像集的人脸/物体识别提出了 IIDLA。该算法由混合卷积层和自适应多层卷积神经网络组成，以非监督的方式预训练 IIDLA 获得初始化参数，并进行特定类特征学习。深度学习算法经过 FOCUT、YT、DMIY 和 ETC-80 等数据集上的人脸/物体识别的评估测试。通过与其他图像识别、分类算法进行比较，结果表明提出的算法性能更优。

此外，IIDLA 在自动视觉识别相关的诸多领域有着巨大的发展潜力。下一步工作中，会使用更复杂的字典学习算法[18]，如稀疏编码和变量代替 K-均值算法，以求进一步扩展。

参考文献：

[1] Lu Yifan, Kang Tingkang, Zhang Huizhong, *et al.* Enhanced hierarchical model of object recognition based on a novel patch selection method in salient regions [J]. IET Computer Vision, 2015, 9 (5): 663-672.

[2] 刘峰，孟凡荣，梁志贞. 基于一阶和二阶信息图像表示的人脸识别 [J]. 计算机应用研究, 2017, 34 (2): 603-606. (Liu Feng, Meng Fanrong, Liang Zhizhen. Face recognition based on first and two order information and image representation [J]. Application Research of Computers, 2017, 34 (2): 603-606.)

[3] 马思伟. AVS 视频编码标准技术回顾及最新进展 [J]. 计算机研究与发展，2015, 52 (1): 27-37. (Ma Siwei. AVS video encoding standard technology review and the latest progress of [J]. Computer research and development, 2015, 52 (1): 27-37.)

[4] Yang Mi, Wang Xue, Liu Wen, *et al.* Joint regularized nearest points for image set based face recognition [J]. Image & Vision Computing, 2017, 58 (C): 47-60.

[5] Hu Huan, Gu Jian. Multi-manifolds discriminative canonical correlation analysis for image set-based face recognition [J]. Cognitive Computation, 2016, 8 (5): 900-909.

[6] Huang Wen, Absil P A, Gallivan K A. A Riemannian symmetric rank-one trust-region method [J]. Mathematical Programming, 2015, 170 (2): 179-216.

[7] Hu H, Gu J. Multi-manifolds Discriminative Canonical Correlation Analysis for Image Set-Based Face Recognition [J]. Cognitive Computation, 2016, 8 (5): 900-909.

[8] 张强，齐兴斌. 利用同伦摄动法的四阶微分方程数值解求解方法 [J]. 湘潭大学自科学报, 2017, 39 (2): 5-8. (Zhang Qiang, Qi Xingbin. Using the homotopy perturbation method to solve the four order differential

equation numerical solution method [J]. Journal of Xiangtan University, 2017, 39 (2): 5-8.)

[9] Shen Xiaobo, Sun Quansen. Orthogonal multiset canonical correlation analysis based on fractional-order and its application in multiple feature extraction and recognition [J]. Neural Processing Letters, 2015, 42 (2): 301-316.

[10] 马建红，张晗，季秋．利用协方差矩阵法表示深度图像集的鲁棒人脸识别 [J]．计算机应用研究, 2016, 33 (12): 3847-3851. (Ma Jianhong, Zhang Han, Ji Qiu. Using the covariance matrix method to express the robust face recognition of depth image set, [J]. Application Research of Computers, 2016, 33 (12): 3847-3851.)

[11] Qiao Hai, Li Chun, Yin Ping, *et al.* Human-inspired motion model of upper-limb with fast response and learning ability–a promising direction for robot system and control [J]. Assembly Automation, 2016, 36 (1): 97-107.

[12] 蔡慧英，朱枫．仿射变换下基于凸包和多尺度积分特征的形状匹配方法 [J]．计算机辅助设计与图形学学报, 2017, 29 (2): 269-278. (Cai Huiying, Zhu Feng. Shape matching method based on convex hull and multiscale integral features in affine transformation [J]. Computer aided design and graphics, 2017, 29 (2): 269-278.)

[13] Lu Jing, Wang Gang, Deng Wu, *et al.* Simultaneous feature and dictionary learning for image set based face recognition [C]// Proc of European Conference on Computer Vision. [S. l.] : Springer, 2014: 265-280.

[14] Hu Heng. Face recognition with image sets using locally grassmannian discriminant analysis [J]. Computational Science and Its Applications, 2015, 12 (2): 37-42.

[15] Miao Yu, Zhang Hui, Metze F. Speaker adaptive training of deep neural network acoustic models using i-vectors [J]. IEEE//ACM Trans on Audio Speech & Language Processing, 2015, 23 (11): 1938-1949.

[16] 孙晓，潘汀，任福继．基于ROI-KNN卷积神经网络的面部表情识别 [J]．自动化学报, 2016, 42 (6): 883-891. (Sun Xiao, Pan Ting, Ren Fuji. Facial expression recognition based on ROI-KNN convolutional neural network, [J]. Automation journal, 2016, 42 (6): 883-891.)

[17] 李燕，王玲．基于肤色和 Haar 方差特征的人脸检测 [J]．计算机工程与科学, 2015, 37 (1): 146-151. (Li Yan, Wang Ling. Face detection based on skin color and Haar variance feature [J]. Computer engineering and science, 2015, 37 (1): 146-151.)

[18] 赵雯，吴小俊．基于鉴别性低秩表示及字典学习的鲁棒人脸识别算法 [J]．计算机应用研究, 2017, 34 (10): 3157-3161. (Zhao Wen, Wu Xiaojun. A robust face recognition algorithm based on discriminative low rank representation and dictionary learning. [J]. Application Research of Computers, 2017, 34 (10): 3157-3161.)